



[12] 发明专利申请公开说明书

[21]申请号 94104792.X

[51]Int.Cl⁵

G06F 15/20

[43]公开日 1995年5月10日

[22]申请日 94.5.7

[30]优先权

[32]93.11.2 [33]JP[31]274604/93

[71]申请人 松下电器产业株式会社

地址 日本大阪府

[72]发明人 郭俊桔

|74||专利代理机构 上海专利商标事务所 代理人 汪 瑜

G06F 3/00

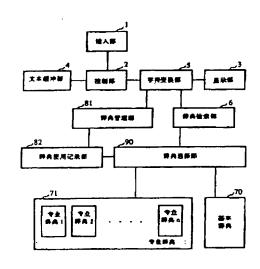
说明书页数:

附图页数:

|54|发明名称 汉字变换装置

[57]摘要

汉字变换装置。其输入部从输入标音字符串截出相应词态繁读音码送至辞典检索部。辞典使用记录部分别对各辞典记录使用次数。辞典选择部让使用次数多的辞典优先检索。辞典检索部以截出读音码为检索码并按优先顺序检出相应的所有同音字、候选词。由文字变换部将它们显示于显示部,供使用者确认正确的变换并把登记有认为正确的字、词的辞典的编码送至辞典管理部。辞典管理部据此更新辞典使用记录部。从而能恰当选择辞典。



(BJ)第 1456 号

- 1.一种汉字变换装置具有将与输入标音字符串对应的词态素读音码截出后送至辞典检索部的输入部、由基本辞典与多本专业群构成的辞典部和显示部,其特征是还具有:①对各辞典记录其使用次数的辞典记录部;②让上述辞典使用记录部所记录的使用次数多的优先来决定各辞典检索顺序的辞典选择部;③以上述输入部送来的读音记号为检索关键码,按上述辞典选择部决定的检索顺序进行各辞典的检索,取出与读音记号对应的所有同音字、候选词以享符变换部的辞典检索部;④使上述辞典检索部检出的同音字、候选词显示于显示部,让使用者确认正确的变换,同时将登记有被认为正确的时典检索部,让使用者确认正确的变换,同时将登记有被认为正确的时等次,让使用者确认正确的方符变换部。⑤根据上述字符变换部送来的辞典编码,按所定程序更新上述辞典使用记录部记录的各辞典的使用次数的辞典管理部。
- 2. 根据权利要求 1 所述的汉字变换装置,其特征是上述辞典管理部更新上述辞典使用记录部的记录时,在记录有使用次数的多本辞典中,规定数量辞典的使用次数达到预定阈值情况下,对迄今的或者汇总汉字变换结束前的使用次数未达到其阈值的辞典、将其使用次数减去规定值而为 0,或者迄今使用次数为 0 而经上述减法结果为负的辞典,设定为不使用。
- 3. 根据权利要求 1 或 2 所述的汉字变换装置,其特征是上述辞典检索部在规定数量辞典中检索到与读音记号对应的同音字、候选词时,上述显示部只将该阶段中检索到的同音字、候选词显示于显示部,以便使用者确认正确的变换。

汉字变换装置

本发明涉及汉字变换装置,特别是关于按照输入文件正确选择 基本辞典和多个专业辞典。

为了用以往技术、实施例等说明本发明,先对重要的单词、用语等进行说明、定义。

本发明的主要对象是中文文字处理器、日文文字处理器等文字变换装置。因此,本发明说明书中所述的由使用者或应用业务程序等所输入的作为汉字变换对象的"标音字符",除罗马字母那样的纯粹标音字符外,当然还包括像日文中假名那样的音素字符,另外还包括阿拉伯数字那样的表意文字的字符,还有朝鲜语中的谚文(朝鲜文字)、中文中的字音记号,例如类似音调记号"—"、"/"、"V"、"\"的记号(这些在电子信息处理系统的输入中以类似记号代用)。而且,作为汉字变换结果的"汉字"不仅包括如"法"、"特许法"的纯粹汉字、较确切地只由"表意文字"、汉字构成的单词,不用说还包括如"爱する"等汉字与假名混和的单词、句节,此外,还包括"々(のま)"等记号。

关于说明中常使用如"标音字符串"等词汇中出现的"串"这样的汉字,由于实际上常常是即使一个汉字也用多个标音字符指定变换对象,而且考虑到变换结果是多个汉字的情况为多,所以不特别限定为复数。即,也包含单个的情况。还有,"语""词"等也往往由独一文字构成。

至于"专业辞典"中的"专业"的概念,可以说是"相互关连的单词、句子等的统称",不仅包括"佛教"、"共产主义"、"法律"、"军事"

"外交"等学术或者实际业务方面的分类,还包括"明代"、"汉代"、"周以前"等时代的划分、"北京话"、"上海话"、"日语"、"朝鲜语"等语言的区分、以及"姓名"、"艺人"、"中文化的外语"等特殊区分。

汉字读音的输入方式,例如通过键盘输入表示汉字读音的标音 字符,对照辞典进行词态素的读音码截出和文字变换处理,再将输 入的标音字符串变换成文字串。然而,在词典中含许多字、语的情况 下, 词态素的读音码的截出容易搞错, 会因同音字、语增加和辞典 检索速度降低而降低汉字变换的效率。因此,一般将辞典分为收存 (在收集基础上, 存贮成可利用状态) 日常的一般性单词的基本辞 典和收存各专业所使用的单词的专业辞典。这种情况下,使用者就 可以根据想要输入的文件的专业选择专业辞典。 可是, 当想输入的 文件跨及几个专业时,使用者需经常转换专业辞典,因此输入效率 变差。以往,具有这种转换专业辞典功能的汉字变换装置有日本特 许公开号平4-80863号公开的装置。图5为这种汉字变换装置一例 的结构图。该图中,1为输入部、2为控制部、3为显示部、4为文本缓 冲器、5 为字符变换部、6 为辞典检索部、70 为基本辞典、71 — 7n 为 专业辞典、8为辞典选择顺序指定部、91及92为辞典群保持部。辞典 群保持部 91、92 按照与使用者编制的文件内容对应的检索顺序保持 基本辞典 70 和多个专业辞典 71-7n。输入部 1 具有输入假名字符 串等、或在假名汉字变换时进行切换辞典群保持部 91、92 的操作的 键盘。文本缓冲器4保持输入标音字符串和变换后的汉字串等。显 示部 3 具有 CRT 等,显示文本缓冲器 4 的内容等。辞典选择、顺序 指定部8根据编制文件的内容并按照其检索顺序从基本辞典70及 多个专业辞典 71-7n 中选择想检索的多个辞典组,进行登记在辞 典群保持部 91、92 上的处理。辞典检索部 6 检索保持在辞典群保持 部 91、92 中的辞典,以便将输入的假名字符串变换成假名汉字混和 字符串。假名汉字变换控制部5根据辞典检索部6检索到的辞典将

输入的假名字符串变换成假名汉字混和字符串。控制部 2 控制输入部 1、文本缓冲器 4、显示部 3、辞典选择、顺序指定部 8、假名汉字变换控制部 5 等。

以下对上述构成的具有专业辞典转换功能的汉字变换装置,说 明其工作。使用者根据编制文件内容,用辞典选择、顺序指定部8从 基本辞典 70 及多个专业辞典 70-7n 中选择将成为检索对象的、汉 字与单词登记可能性高的多个辞典,并再指定其检索顺序,将其结 果登记于辞典群保持部 91、92。例如,编制有关信息处理文件时,把 通常变换时想检索的辞典 (基本辞典及例如对应文件内容的信息处 理用语辞典)、欲变换为固有名词等时把想检索的辞典(例如,姓名 辞典、地名辞典、基本辞典)分别以其优先顺序预先登记在辞典群保 持部 91、辞典群保持部 92 中。由输入部 1 输入的假名字符串保持在 文本缓冲器 4 内并由显示部 3 显示。假名汉字处理中,假名汉字变 换控制部5一边用辞典检索部6检索登记于辞典群保持部91、92中 的辞典,一边将假名字符串变换成假名汉字混和的字符串。进而用 变换后的字符串置换文本缓冲器 4 内的假名字符串,显示在显示部 3上。此时,在预先将上述辞典登记在辞典群保持部91、92中的情况 下,通常的假名汉字变换时,因按登记顺序检索登记在辞典群保持 部 91 中的辞典, 信息处理用语比一般用语先输出。若文件中出现人 名等固有名词,则操作输入部1的辞典群转换键,使被检索的辞典 群转换至登记在辞典群保持部92中的辞典。若再次操作辞典的转换 键,则回到辞典群保持部91。就这样通过操作辞典群转换键来进行 辞典群的转换。

另外,因有些内容与本发明宗旨无直接关系,省去有关图示与说明。例如同一发音下有多个汉字串、词语的情况下用于将它们按优先顺序或一次全显示于屏幕后让使用者特别指定想要的字语的程序及其所必要的结构;用于将使用者特别指定的汉字串显示于眼下正

写作中的文章的所定位置的程序及结构;用于将由许多输入标音字符组成的标音字符串划分成作为整体的汉字变换对象的必要程序,例如最长一致法及实现其的结构,与通常的中文文字处理器、日文文字处理器相同或基本相同。

然而,上述以往技术装置存在以下问题。

第一个问题:由工作人员按输入文件内容选择辞典,且必需转换分专业的辞典。而正确转换专业辞典,对于使用者来说成为大负担。还有,登记在辞典群保持部内的辞典因专业领域不正确时,由最长一致法等截出的作为汉字变换对象的词态素读音码(标音字符串)以及文字变换就不正确,汉字变换效率变差。

第二问题: 若输入文件中包括多个专业的单词,为此而必需对专业选定作繁琐变更时,对使用者来说,转换专业辞典很费时间。

本发明的目的在于解决上述问题,提供一种能自动又确切选定专业辞典的汉字变换装置。

 的辞典管理部。

本发明第二形态的汉字变换装置,是在第一形态基础上再加上这样的特征,即上述辞典管理部更新上述辞典使用记录部的记录时,在记录有使用次数的多本辞典中,规定数量辞典的使用次数达到预定阈值情况下,对迄今的或者汇总汉字变换结束前的使用次数未达到其阈值的辞典、将其使用次数减去规定值而为 0,或者迄今使用次数为 0 而经上述减法结果为负的辞典,设定为不使用。

本发明第三形态的汉字变换装置,是在第一或二形态基础上再加上这样的特征,即上述辞典检索部在规定数量辞典中检索到与读音记号对应的同音字、候选词时,上述显示部只将该阶段中检索到的同音字、候选词显示于显示部,以便使用者确认正确的变换。

- 图 1 是本发明汉字变换装置一实施例的结构图;
- 图 2 是图 1 实施例的处理流程图;
- 图 3 是上述实施例的辞典管理部的工作流程图;
- 图 4 从概念上表示图 1 实施例的辞典使用记录部的内容;
- 图 5 是以往汉字变换装置的结构图。

图中有关标号含义为以下。1:输入部、2:控制部、3:显示部、4:文本缓冲部、5:字符变换部、6:辞典检索部、70:基本辞典、71·····7n:专业辞典、81:辞典管理部、82:辞典使用记录部、90:辞典选择部。

以下举实施列说明本发明。

(第1实施例)

本实施例是关于本发明第一形态和第二形态的发明。

图1是本发明汉字变换装置一实施例的结构图。图中,1是输入部、2是控制部、3是显示部、4是文本缓冲部、5是文字变换部、6是辞典检索部、70是基本辞典、71-7n是专业辞典、81是辞典管理部、82是辞典使用记录部、90是辞典选择部。控制部 2、显示部 3、文本缓冲部 4、基本辞典 70及专业辞典 71-7n 的结构、原理与以往技术相

同或没有大的区别,这方面的说明只要与本发明无关,就省略或是 简单说明。基本辞典70收存了以读者记号为检索关键码的对应的基 本单词。辞典使用记录部81以辞典编码为检索关键码,记录使用次 数值或不使用记号"X"的值,而且按其使用次数值大的比小的排前 的顺序进行记录。专业辞典 71-7n,对多个专业领域收集各专业固 有的单词,并以读音记号为检索关键码收存该专业中对应的单词。 辞典选择部90参考辞典使用记录部82所记录的各专业辞典的使用 次数及排列顺序,决定使用的专业辞典及其检索顺序。还根据辞典 管理部 81 的指示,决定作为检索对象的专业辞典和检索顺序。辞典 检索部 6,以读音记号为检索关键码,根据辞典选择部 90 指定的专 业辞典及其检索顺序,取出对应的所有同音字、候选词。辞典管理部 81,对每个由使用者确认为正确变换的字符串中的汉字、单词,根 据它们的辞典编码更新处理辞典使用记录部82所记录的各辞典的 使用次数记录及排列顺序。字符变换部5,通过辞典检索部6将输入 标音字符串变换成候选变换的汉字串、单词后,使之显示于显示部3 上,然后,让使用者选择想要的汉字串、单词,同时将登记有确认为 正确的汉字串、单词的辞典的编码送至辞典管理部81。显示部3依 次显示辞典检索部 61 检索过的同音字、候选词,进而将使用者认为 正确的汉字串、单词显示于写作中的文章的所定位置。

图 2 和图 3 是本实施例的工作流程图。以下根据图 2 说明其基本的工作流程。

(S21):使用者通过输入部将作为汇集一起的汉字变换对象的标音字符串进行输入。并在后文将举的具体例中,输入构成7个音节的标音字符串。

(S22):输入部从输入标音字符串中截出与成为一体变换对象的单词对应的词态素读音码。并在后文将举的具体例中,2个或者3个音节构成单词。

(S23):辞典检索部参照辞典使用记录部记录的使用次数,检索专业辞典或基本辞典的内容,显示其对应的字、候选词。此时,若是检索被认为是未变换或错变换的词态素的全部读音码,则对该词态素进行字符变换。图 4 从概念上表示该辞典使用记录部的使用次数的记录内容。

(S24):使用者参见显示部,判断被变换的汉字是否全都正确。若有不正确的汉字,则应将它指出,并参照该词态素或者是包括其在内的成为一体变换对象的词态素的读音码再次进行检索,回到(S23)。若没有,则进入(S25)。

(S25):参照包含被认为是正确对应词态素读音码的汉字、单词的辞典的辞典编码,更新辞典使用记录部的内容。后文将详细说明其更新时的详细工作过程。

(S26):将正确变换的汉字串显示于写作中的文章的所定位置上。

下面,根据图3说明辞典管理部的工作过程。

(S301):设定阈值 TH 以及作为汇总的汉字变换对象而输入的标音字符串中所含的成一体性汉字变换的单词的总数 h。

(S302): 将与确定为正确变换的单词对应的汉字、单词的辞典编码依次存入暂存装置。为此,设定变量 A[J]。其中, J=1、2、…… h, 各 A[J]的内容为辞典编码。

(S303):设变量J的起始值为1。

(S304): 判断是否 J>h。若为否定,则进入(S305)。

(S305):以A[J]之值为检索关键码,参见辞典使用记录部,取出该辞典的使用次数,作为变量 CNT 进行存贮。

又因 CNT=X 情况下,该辞典为不使用,所以不会化费太多的检索时间。

(S306):CNT 加 1 后,就进入上述(S307)。

(S307):判断是否 CNT>TH。若为肯定,则为了转成更新处理,进入(S308),若为否定,则进入(S310)。

(S308): 让当前辞典使用记录部所记录的辞典的使用次数减去规定值 D。然后,进入(S309)。

(S309): 将此结果设定为新使用次数。使用次数为 0 或者为负数的辞典,将其使用次数设定为 X。进入(S310)。

(S310):以A[J]为检索关键码求该辞典,并将该新的 CNT 之值记录于辞典使用记录部。

(S311):J加上1后,返回至应以下一个汉字、单词为对象而进行相同处理的(S304)。

(S312):根据使用次数,结束辞典使用记录部的使用次数的记录和排列顺序的更新。

图 4 从概念上表示辞典使用存贮部的存贮内容。本图中,第 1 列表示专业辞典编码,第 2 列表示使用次数,将它们按使用次数大的顺序从上至下进行排列。图中,使用次数为 X 的表示该专业辞典定为不使用。

本实施例中,基本辞典除作为本发明的作用对象外,还被列为第一优先检索,将辞典管理部的阈值和比较专业辞典使用次数的规定数设定为1。

实际上,因基本辞典的使用频度比任何一个专业辞典高,这与将基本辞典也作为本发明的作用对象后设置阈值与比较使用次数的辞典的规定值为2实质上是相同的。

但是,由于此情况下只是基本辞典的使用频度变高,往往总是其使用次数的记录数变高,因此有必要采取设置其上限等措施。

接着,举图4所示辞典使用存贮部为例,按图3所示辞典管理部的工作过程,具体说明其如何变化。

现若输入"lian2jyuen1zuen1shou3guo2ji4fa3"这样的标音字符

串,则根据使构成音节数第1优先、使输入顺序第2优先而截出作为汉字变换对象的音节的最长一致法截出词态素的读音码,结果为"lian2jyuen1 zuen1shou3 guo2ji4fa3"。然后,参见专业辞典和基本辞典,将上述标音字符串变换成"联军遵守国际法"这样的字符串。其关连的辞典编码分别是"联军"为5(军事)、"遵导"为0(基本辞典)、"国际法"为7(法律)。又因为在前面图3所示工作过程说明中,各A[J]的内容只是"专业辞典的编码",专业辞典的种类全部是9个,因此(S302)中的A[J]之值为1~9。其中,辞典编码为"0"的表示基本辞典。

接着,说明图3的辞典管理部的处理过程。

(S301):由于上述一连串标音字符串的变迁中,成为一体性汉字变换对象的汉字串、单词的个数为 3,故确定 h=3。为便于说明,在此将 TH 的值设定为 137。

(S302): A[1]=5, A[2]=0, A[3]=7

(S303): 设变量 J 为 1。

(S304): 因 J < 3 = h, 进入(S305)。

(S305):以A[1]之值 5 为检索关键码,从辞典使用记录部取出其使用次数 137 后存入 CNT 中。

(S306):CNT 加上 1,成为 138。

(S307): 因 CNT>TH, 进入(S309)。

(S308): 辞典使用记录部由于减去规定值 10, 更新成以下那样。(但只表示图 4 中的更新部分。而且,表示以后更新内容时也相同。)

5(军事) 138

3(经济) 59

7(法律) 1

8(会计) X

在以上条件下,进入(S310)。

(S310):以A[1]之值 5 为检索关键码求出辞典,并将当前的 CNT 之值作为使用次数,记录于辞典使用记录部。

(S311): J加上1为2后,返回上述(S304)。

J=2、J=3 中也进行相同处理工作。于是,此状态下的辞典使用记录部的内容依次表示为如下。

J=2:

5(军事) 139

3(经济) 59

7(法律) 1

8(会计) X

J = 3:

5(军事) 139

3(经济) 59

7(法律) 2

8(会计) X

此时,因(304)中J=4>"3",进入(S312)。

(S312): 结束使用次数的记录及与其对应的排列顺序的变更处理。此时, 最后使用的专业辞典与其排列顺序相同。

(第2实施例)

本实施例,除了在前述第1实施例中,当某专业辞典的使用次数达到预定的阈值时,辞典管理部将达到此情况下的使用次数改设为0以外,都与第1实施例相同。

因此,省去的各部结构图的图示以及以此为基础的工作作用说明等,并省去处理过程图中相同的工作说明,只说明本实施例固有的工作,处理过程等。

在前述的(S307)中,辞典管理部将有的专业辞典之值设为 0。

于是, 辞典使用记录部的内容为以下那样。

- 5(军事) 0
- 3(经济) 0
- 7(法律) 0
- 8(会计) 1

同样,在(S311)中,于J=Z的工作结束阶段,辞典使用记录部的内容为以下那样。

- 5(军事) 1
- 3(经济) 0
- 7(法律) 0
- 8(会计) X

在 J=3 的工作结束阶段,同样为以下那样。

- 5(军事) 1
- 3(经济) 0
- 7(法律) 1
- 8(会计) X

在上述使用次数下,辞典选择部变更各专业辞典的排列顺序。因此,最后为下述那样。

- 5(军事) 1
- 7(经济) 1
- 3(法律) 0
- 8(会计) X

与前面第1实施例相比,因本实施例将所有的专业辞典的使用次数改设为0,这对作为起初使用的专业辞典设定数(种类)为多的情况变得有利。

(第3实施例)

本实施例是关于本发明的第三形态。

本实施例,因基本结构与前述第1实施例相同,省去结构图等。

不同之处只是辞典选择部内装计数部,该计数部每逢一检索关键码即读音记号就计算作为检索对象的辞典的个数,若该计算值在预定的规定值内并检索成功,则将该情况通知字符变换部。而且,接到该通知的字符变换部使该阶段检索到的同音字、候选词显示于显示部的所定位置,以便使用者确认正确的选择。并且,具有将通过检索其后优先序位低的辞典检出的同音词、候选词不显示于显示部而进行保存的暂存部。

至于计数部、根据来自计数部的通知使字符变换部在中途阶段 发挥作用的扳机信号(触发信号)发生部、暂存部的结构都因是众所 周知的技术,省去其内容说明。

与前面实施例相比,本实施例在检索中途阶段单把正确率高的同音字、候选词作为使用者想要的汉字、单词的确认对象而进行显示、选择,从而提高了汉字变换速度。

以上根据实施例说明了本实施例,不用说,本发明不只限于上 述实施例。也就是说,例如以下情况也包含于本发明内。

- (1)第1、第2实施例中,是加X表示不使用的专业辞典,但也可以采用其它任意标记、措施。同样,各专业辞典的使用次数减去规定值,不是逐个单词(具体例中的A[J],对J=1、2、3逐个地)进行,而是对作为汇总的汉字变换对象的全部单词的各专业辞典添加修正其使用次数后(具体例中,对"联军"、"遵守"、"国际法"全进行修正后),依次比较阈值,只要一专业辞典其使用次数达阈值,就进行减去规定值。
- (2)通过使用控制寄存器累加各辞典每次更新次数,当其值大 于某阈值时也更新辞典使用记录部。
- (3)在本发明各形态中,基本辞典除作为辞典使用记录部、辞典选择部、辞典管理部、辞典检索部的作用对象外或者除按使用者意思

成为作用对象外还能添加将某专业辞典指定为基本辞典的功能。

同样,使用者对特定的专业辞典除从起初就成为作用对象以外,还能添加其它可能的功能。此外,也可增加可变更本发明第二形态的规定数、阈值、规定值的功能。

- (4)添加能装、卸专业辞典的功能。
- (5)字符变换部让辞典检索部检索到的同音字等显示于显示部时,添加在由于检索到许多同音字而难以直接全部显示的情况下从使用频度高的开始进行显示的功能,或通过让使用频度高的显示得大,低的显示得小而一次性进行显示等功能。
 - (6)对基本辞典、各专业辞典分别添加可发挥学习能力的装置。
- (7)将辞典使用记录部的记录内容本身记录在其他途径的装置上并可根据需要而读出、转换。因此,能根据写作文件的专业情况,将以前写过的高一专业文件的使用经验反映至现在写作中的文件中,而且能在同一文字处理机等上有效作成多专业的文件。
- (8)为便于制造,可将本发明所述形态中的一构件要素(构成要件)在物理和机械方面分为多个构件,反之也可将多个构件做成一个构件,或者将它们适当组合。
- (9)在检不出同音字、候选词时,添加选择机构,以便将此情况向使用者显示并自动发挥作用检索被定为不使用的专业辞典。

综上所述,若利用本发明,则具有以下效果。

- (1)不必由人工选择专业辞典,对于包含许多专业的单词、汉字的文件写作,汉字变换效率高。
- (2)由于能获得以专业辞典为对象的使用经验方面的学习能力,通过优先选择必要的专业辞典、关闭不必要辞典,提高辞典的检索速度。
- (3)由于平时只保持必要的专业辞典,作为汉字变换的候选而检索及列举的同音汉字、候选词也变少。因此不仅大为提高字符变换

正确率和速度,而且还减轻使用者的用眼疲劳等。

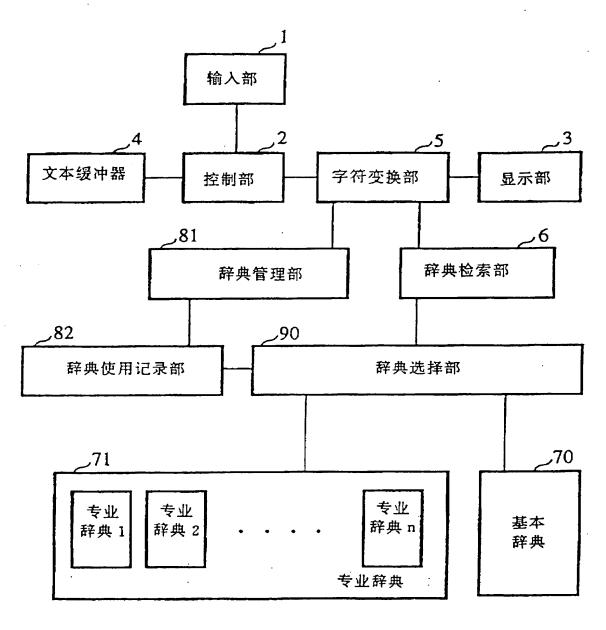
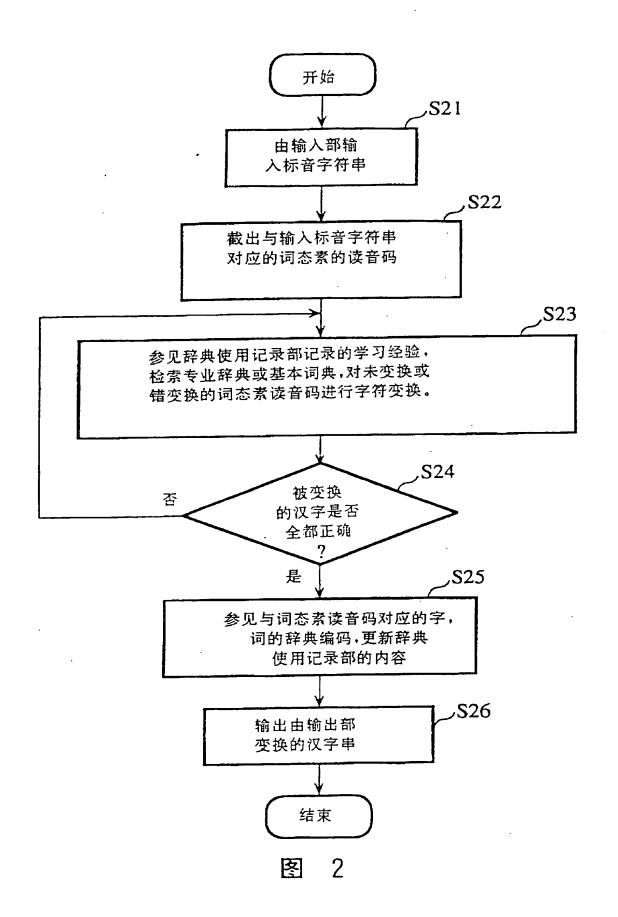
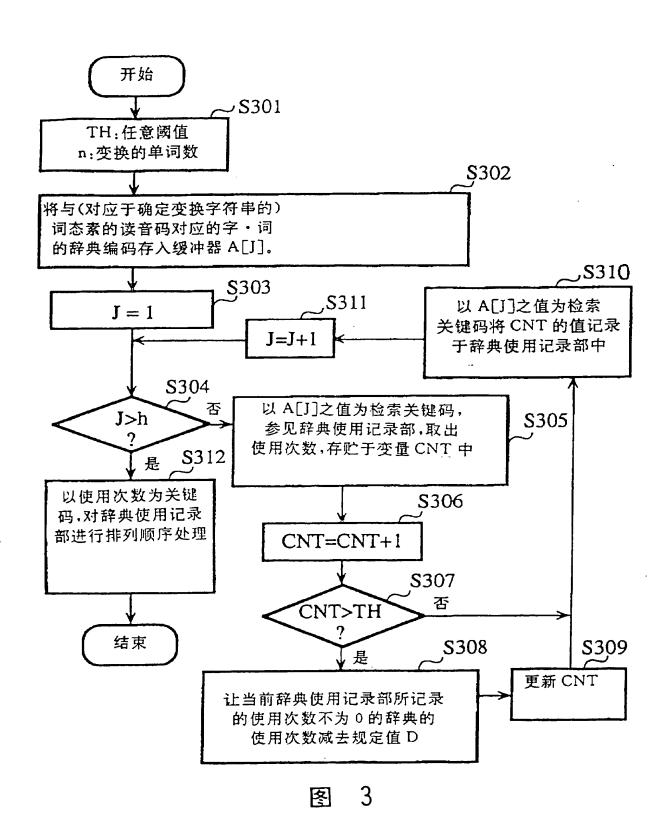


图 1



8



- 3 -

专业辞典编码

使用次数

5 (军事)

137

3 (经济)

69

7 (法律)

11

8 (会計)

1

1 (教育)

X

9 (医学)

 \mathbf{X}

4 (机械)

X

2 (政治)

X

6 (生物)

X

(X:不使用)

图 4

